

November 10, 2021

To: James Tucker

Chair

National Advisory Committee

From: Ron S. Jarmin

Acting Director U.S. Census Bureau

Subject: Recommendations and Comments to the U.S. Census Bureau from the National

Advisory Committee 2021 Differential Privacy Meeting

The U.S. Census Bureau thanks the National Advisory Committee for its recommendations. We are responding to the committee recommendations from the 2021 Differential Privacy Meeting on May 27, 2021.

Your feedback is welcomed to ensure that the Census Bureau continues to provide relevant and timely statistics used by federal, state, and local governments, as well as business and industry, in an increasingly technologically oriented society.

Attachment



To: Ron S. Jarmin

Acting Director
U.S. Census Bureau

From: James Tucker

National Advisory Committee (NAC) Chair

Subject: Recommendations and Comments to the Census Bureau from the National

Advisory Committee – Differential Privacy Meeting

May 27, 2021

I. Priority use cases

Much of the focus on the Census Bureau's application of its Disclosure Avoidance System (DAS) on the P.L. 94-171 redistricting data set (P.L. Data) has, unsurprisingly, been on redistricting. However, that focus overlooks the many other priority uses for the P.L. data.

The P.L. Data is used for appropriation decisions where the highest level of accuracy is needed. Even the loss of a small number of people from a community's population reported in the P.L. Data following application of the DAS can significantly impact that community.

Differences that may be perceived as "minor" and falling within an "acceptable level" of variance from the actual 2020 Census count through application of the DAS can have dire consequences for less populated tribal communities and places.

- 1. **RECOMMENDATION:** The NAC recommends that in evaluating the application of the DAS to the P.L. 94-171 redistricting data set, including the algorithm and the Privacy Loss Budget (PLB), the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee consider priority uses that require the highest level of accuracy, including:
 - a. Redistricting of all elected bodies, including those with very small populations;
 - b. Funding formulas and appropriations, including for communities with very small populations;
 - c. Research and public health, including for communities with very small populations;
 - d. Tribal, state, county and local level governance, including in places with very small populations.

The Census Bureau has acknowledged that the greatest level of variance between the actual count in the 2020 Census and the adjusted count following application of the DAS to the P.L. Data occurs in places with 500 people or less. According to the Census Bureau, about 76% of the more than 19,500 incorporated places have fewer than 5,000 people and about a third of those have less than 500 people. That means that approximately 5,000 places in the United States, or roughly one quarter of all incorporated places, are likely to face the greatest impact of variance caused by noise added through application of the DAS. Those places represent tribal communities and small towns that depend upon an accurate count to secure much-needed resources.

For example, if the actual count from a community was 400 people and application of the DAS results in a decrease of that count of just 5 percent, or 20 people, that community could lose between \$60,000 and \$100,000 a year in federal appropriations each year for ten years (depending upon the particular programs that are included). That could result in the community having to cut one or two full-time employees who provide vital social services to that community.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation. The 2020 Disclosure Avoidance System (DAS) production settings approved by the Data Stewardship Executive Policy (DSEP) Committee for the production run of the P.L. 94-171 Redistricting Data Summary File were selected based on a thorough assessment of accuracy and fitness-for-use for a wide range of priority-use cases, including those identified by the committee.

2. **RECOMMENDATION:** The NAC recommends that in evaluating the application of the DAS to the P.L. 94-171 redistricting data set, including the algorithm and the Privacy Loss Budget (PLB), the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee consider the impact that even small levels of inaccuracy can have on tribal communities and places with 500 people or less.

Census Bureau Response: The Census Bureau accepts this recommendation. In evaluating the application of the 2020 Disclosure Avoidance System (DAS) to the P.L. 94-171 redistricting data set, the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee considered the impact on tribal communities and places with 500 people or less.

3. RECOMMENDATION: The NAC recommends that in evaluating the application of the DAS to the P.L. 94-171 redistricting data set, including the algorithm and the Privacy Loss Budget (PLB), the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee consider the level of variance between the actual count and the adjusted count at the Census Block level.

Census Bureau Response: The Census Bureau accepts this recommendation. When making its decisions regarding the production settings for the P.L. 94-171 Redistricting Data Summary File, the Data Stewardship Executive Policy (DSEP) Committee closely examined the amount and distribution of noise in block level statistics to ensure that it would be sufficient to protect respondent confidentiality, consistent with the legal requirements of Title 13 sections 8 and 9, while also ensuring the reliability and validity of the data when individual blocks are aggregated into user-defined geographies for the purposes of redistricting and Voting Rights Act enforcement, among other uses.

The NAC applauds the Census Bureau for the success of its Voting District Program, which has allowed state, county and local jurisdictions to update the Census geography to match up to the geography used for voting districts or precincts. Using the same geography for Census data and precinct-level political data is important because it facilitates drawing districts that provide racial and ethnic minorities equal opportunities to elect their candidates of choice, as provided by the federal Voting Rights Act.

The NAC is concerned that application of the DAS may inject distortions into the demographic data used from the P.L. Data so that it no longer accurately correlates with precinct-level data.

4. RECOMMENDATION: The NAC recommends that in evaluating the application of the DAS to the P.L. 94-171 redistricting data set, including the algorithm and the Privacy Loss Budget (PLB), the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee consider the level of variance between the actual count and the adjusted count for small political boundaries added through the Voting District Program, such as voting districts or precincts.

Census Bureau Response: The Census Bureau accepts this recommendation. In evaluating the application of the 2020 Disclosure Avoidance System (DAS) to the P.L. 94-171 redistricting data set, the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee considered the level of variance between the actual count and the confidentiality-protected count for small political boundaries added through the Voting District Program.

To improve the accuracy of data for more sparsely populated places, the Census Bureau has prioritized obtaining a more accurate count for the largest minority group in that place. However, in many cases that adjustment comes at the expense of smaller minority groups. In particular, analysis of the latest demonstration data has identified that in many small places, the smallest minority groups lose most, and in some cases all, of their population. Those adjustments have profound implications for the smallest minority groups being sacrificed to secure a higher level of accuracy for the largest minority group.

5. RECOMMENDATION: The NAC recommends that in evaluating the application of the DAS to the P.L. 94-171 redistricting data set, including the algorithm and the Privacy Loss Budget (PLB), the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee consider the level of variance between the actual count and the adjusted count for all minority groups in less populated places, and not just the largest minority group.

Census Bureau Response: The Census Bureau accepts this recommendation. In evaluating the application of the 2020 Disclosure Avoidance System (DAS) to the P.L. 94-171 redistricting data set, the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee considered the level of variance between the actual count and the confidentiality-protected count for all minority groups in less-populated places.

The NAC applauds the Census Bureau for its efforts to engage with outside stakeholders to identify priority use cases for the P.L. Data.

6. RECOMMENDATION: The NAC recommends that the Census Bureau continue to work with federal, state and local agencies, Tribal Nations, and other outside stakeholders to collect and study cases across a wide range of uses, variables, and geographies for the P.L. 94-171 redistricting data set and for subsequent data releases.

Census Bureau Response: The Census Bureau accepts this recommendation. We will continue our stakeholder outreach and engagement activities to ensure that our data users' needs are reflected in future decision-making for the Demographic and Housing Characteristics file and subsequent data products.

Other recommendations to improve accuracy and transparency

The Census Bureau has not provided the public with a concrete measure for the loss of privacy if epsilon is adjusted to improve accuracy. What would moving from an Epsilon of 12.3 to 15 mean in terms of loss of privacy protection? This information is needs to be provided to establish view of where to strike the balance between accuracy and privacy.

7. RECOMMENDATION: The NAC recommends that the Census Bureau provide the public with a concrete measure for the loss of privacy associated with adjustments to Epsilon to improve accuracy, so that more informed decisions can be made about the relative tradeoffs between accuracy and privacy.

Census Bureau Response: The Census Bureau accepts this recommendation. The privacy-loss budget approved by DSEP for the production run of the 2020 Census P.L. 94-171 Redistricting Data Summary File is ε =17.14 for the persons file and ε =2.47 for the units file, which represent an exponential increase (more accuracy) over the levels used for the April 2021 Demonstration Data (which were ε =10.3 and ε =1.9, respectively). We are

currently evaluating how to best convey this information to our data users and will keep the committee informed of our efforts in this regard.

The NAC joins NCAI and other organizations in emphasizing to the Census Bureau the importance of the federal government's trust responsibility to ensure that Tribal Nations are counted accurately. The total number of AIAN geographies with zero counts must be zero. If an AIAN geography or other geography had at least one person counted in the 2020 enumeration by any method (whether self-response, proxy, administrative records, imputation or otherwise), they should have counts in the 2020 Census data.

8. RECOMMENDATION: The NAC joins NCAI and other organizations in recommending that the Census Bureau ensure that AIAN Tribal Nations and other ethnic and racial population groups that have actual counts in the 2020 Census raw data never end up with zero counts in the final dataset.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but must reject this recommendation as written. The only mechanism by which the 2020 Census Disclosure Avoidance System could absolutely guarantee that a geographic area or demographic group with a non-zero population would not be reported as zero in the privacy-protected data would be to impose an invariant on the population count for that geographic area or demographic group. The list of invariants for the production run of the 2020 Census P.L. 94-171 Redistricting Data Summary File was approved by the Data Stewardship Executive Policy (DSEP) Committee on November 24, 2020. The DSEP Committee did, however, evaluate this possibility and sought to minimize the likelihood that these types of distortions would occur.

9. RECOMMENDATION: The NAC joins NCAI and other organizations in recommending that consideration be given for the priority use cases listed above, and the Census Bureau must demonstrate how they have updated the application of privacy protections to ensure that AIAN Tribal Nations and other ethnic and racial population groups and groups working on behalf of children have accurate and usable data for all of those purposes.

Census Bureau Response: The Census Bureau accepts this recommendation. In evaluating the application of the 2020 Disclosure Avoidance System (DAS) to the P.L. 94-171 redistricting data summary file, the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee considered the priority use cases listed above, including the accuracy of data for AIAN tribal nations, racial and ethnic groups, and children.

10. RECOMMENDATION: The NAC joins NCAI in recommending that the allocation of the Privacy Loss Budget (PLB) be increased for AIAN Tribal Nations, but not at the expense of other groups, including smaller minority populations.

Census Bureau Response: The Census Bureau accepts this recommendation. The 2020 Disclosure Avoidance System (DAS) production settings approved by the Data Stewardship Executive Policy (DSEP) Committee for the production run of the P.L. 94-171 Redistricting Data Summary File include a higher overall privacy-loss budget and improvements to the DAS post-processing geographic hierarchy.

11. RECOMMENDATION: The NAC joins NCAI and other organizations in recommending that the global Privacy Loss Budget be increased for the P.L. 94-171 redistricting data set.

Census Bureau Response: The Census Bureau accepts this recommendation. The privacy-loss budget approved by DSEP for the production run of the 2020 Census P.L. 94-171 Redistricting Data Summary File is ε =17.14 for the persons file and ε =2.47 for the units file, which represent an exponential increase (more accuracy) over the levels used for the April 2021 Demonstration Data (which were ε =10.3 and ε =1.9, respectively).

The NAC joins NCAI, the National Urban League and other organizations in recommending that the Census Bureau ensure accurate data for all, especially underrepresented racial/ethnic groups and small, rural, and remote groups. Analysis of the metrics reveals a systematic negative impact on these groups. Regardless of the need to protect privacy, the application of the privacy protections and the resulting usability of 2020 Census data must be equitable.

12. RECOMMENDATION: The NAC joins NCAI, the National Urban League and other organizations in recommending that the accuracy for all of racial and ethnic groups match that of non-Hispanic White data in the 2020 Census to be fair and equitable.

Census Bureau Response: The Census Bureau accepts this recommendation. The level and allocation of privacy-loss budget in the 2020 Disclosure Avoidance System (DAS) production settings approved by the Data Stewardship Executive Policy (DSEP) Committee for the production run of the P.L. 94-171 Redistricting Data Summary File reflect adjustments to the settings used in the April 2021 Demonstration Data. The accuracy for all of racial and ethnic groups match that of non-Hispanic White data in the 2020 Census.

13. RECOMMENDATION: The NAC joins NCAI in recommending that further consultation with Tribal Nations and other stakeholders occur after the DSEP decision in June 2021 regarding the algorithm and Privacy Loss Budget to be applied to the P.L. 94-171 redistricting data set. There was feedback from many who have participated in prior consultations that they were not meaningful because the information was too complicated for an audience level unfamiliar with the DAS.

Census Bureau Response: The Census Bureau accepts this recommendation. We are currently evaluating how to better communicate information about the 2020 Census Disclosure Avoidance System to less technical audiences and will schedule additional tribal consultations and other stakeholder engagement opportunities when appropriate.

14. RECOMMENDATION: The NAC joins NCAI and other organizations in recommending that the Census Bureau conduct additional public meetings and Tribal consultations to obtain feedback from Tribal Nations and stakeholders from other ethnic and racial groups and groups working on behalf of children about the decisions made by DSEP and its implications, with information to be communicated in an accessible format for individuals who may lack an understanding for any of the methods the Bureau is using to balance privacy with accuracy.

Census Bureau Response: The Census Bureau accepts this recommendation. We are currently evaluating how to better communicate information about the 2020 Census Disclosure Avoidance System to less technical audiences and will schedule additional tribal consultations and other stakeholder engagement opportunities when appropriate.

15. RECOMMENDATION: The NAC recommends that the Bureau revisit and abandon its policy decision to ensure consistency between redistricting data release and DHC data release, if doing so would secure a greater level of accuracy in DAS-protected data for either release.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but is unable to accept or reject the recommendation at this time. We are currently evaluating the advantages and disadvantages of enforcing or relaxing the consistency requirement and will report back to the committee when that evaluation is complete.

16. RECOMMENDATION: The NAC recommends that the Bureau publicly release the identity and characteristics of all voting/electoral districts evaluated in the course of developing the DP algorithm and epsilon. Characteristics should include urban/rural, voting-age proportion, citizenship proportion, race/ethnicity proportions, total population size, and socioeconomic diversity.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but must reject it at this time. The list of jurisdictions submitted to the Census Bureau for evaluation is currently designated as Controlled Unclassified Information for U.S. Government use only by the U.S. Department of Justice (its owner) and as such cannot be released to the public.

17. RECOMMENDATION: The NAC recommends that the Bureau publicly release data demonstrating how previous DAS (swapping) in 2010 may have affected geographies/ electoral districts that were majority-minority appearing not to be majority-minority.

Definition of majority-minority should employ population percentages identified by the DOJ as generally identifying effective majority-minority districts for each specific minority group.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but must reject it at this time. Many legacy disclosure avoidance methods (including the swapping algorithms used for the 2000 and 2010 censuses) require that their parameters and impacts on the resulting data be kept confidential to protect against reverse-engineering of the original unprotected, confidential data. Consequently, the confidentiality requirements of Title 13 sections 8(b) and 9 prevent the Census Bureau from releasing information on the impact of swapping on fitness-for-use of the published 2010 Census data.

The NAC is particularly concerned that southern Black population (where as much as a third of the Black population resides) and were hit hard by hurricanes and not well counted, will be adversely impacted by DP application and increased disclosure avoidance measures. The "Privacy Loss Budget" should not reflect or translate into a loss of Constitutional rights, including voting--one person, one vote.

18. RECOMMENDATION: The NAC recommends that the Census Bureau consider impacts the application of DP will have on already undercounted populations affected heavily by the differential undercount, including but not limited to Black men and young Black children, the Hispanic population, the AIAN population, the NHPI population and other hard-to-count populations.

Census Bureau Response: The Census Bureau accepts this recommendation. In evaluating the application of the 2020 Disclosure Avoidance System (DAS) to the P.L. 94-171 redistricting data summary file, the Census Bureau and the Data Stewardship Executive Policy (DSEP) Committee considered the impacts of the DAS on populations that have historically experienced a differential net undercount.

General Communications and Engagement about Census Bureau Products and DP

19. RECOMMENDATION: The NAC recommends that the Census Bureau continue its efforts to regularly communicate updates and engage various stakeholders and user groups during the decision-making process for setting the Privacy Loss Budget (PLB) and its allocation.

Census Bureau Response: The Census Bureau accepts this recommendation. Between April and July 1, we conducted eight educational webinars to help data users better understand the proposed changes and their impact on the data. We have released

multiple newsletters detailing DAS developments and the results of research into data accuracy and alternative disclosure avoidance methods. We will continue and expand this level of engagements as we develop the next sets of 2020 Census data products.

20. RECOMMENDATION: The NAC recommends that the Census Bureau be transparent and actively engage with all stakeholders and user groups as it makes its final decisions about the algorithm used for the P.L. 94-171 redistricting data set and going forward with all remaining products.

Census Bureau Response: The Census Bureau accepts this recommendation. The final decisions on the algorithm were made after much candid feedback from data users and outside experts.

21. RECOMMENDATION: The NAC recommends that the Census Bureau communicate the factors used by the Data Stewardship Executive Policy (DSEP) Committee to set the PLB ("level of epsilon").

Census Bureau Response: The Census Bureau accepts this recommendation. We described these factors in our June 9 newsletter and July 1 webinar.

22. RECOMMENDATION: The NAC recommends that the Census Bureau publish 2020 Census data handbooks for data users targeted to different audiences (AIAN, federal agencies, data for rural areas, media, local government officials, etc.) that parallel the handbooks created for the American Community Survey.

Census Bureau Response: The Census Bureau accepts this recommendation. We are working with the Population Reference Bureau (PRB) on the development of a 2020 Census handbook series explaining differential privacy implementation and its impacts. The first handbook is scheduled for release in early November.

Impact on Children

Dr. William O'Hare's research on the Census Bureau's April 28th, 2021 demonstration product looks at the implications of differential privacy for children. O'Hare found that the injected error (Epsilon 12.3) had little impact on accuracy for large geographic units like states or large counties, but the story is different for smaller places and population groups. For example, the analysis shows that many smaller areas have high levels of error. For example, after differential privacy is applied, 8% of Unified School Districts had child counts with an absolute percent error of 5% or more. Data also showed that 66% of Unified School Districts had absolute numeric errors of 10 or more children. Percentagewise the errors are worse for Black, Hispanic, and Asian children. Errors of this magnitude can impact federal and state funding that schools receive, as well as potentially affect funding allocations from the US Department of Health and Human

Services and US Department of Agriculture. It is a problem if communities lose federal funds to support children because of the introduction of artificial error and not because of their changing population.

23. RECOMMENDATION: The NAC recommends that the Census Bureau supply differential privacy metrics for children to better understand how differential privacy will impact data quality.

Census Bureau Response: The Census Bureau accepts this recommendation. We will include additional accuracy metrics for children in future demonstration data releases.

24. RECOMMENDATION: The NAC recommends that the Census Bureau establish the following accuracy target for differential privacy and children: "For all geographic units larger than census blocks for which data are reported, after the census data has been subjected to DP, the results for children should be within 5% of the true value 95% of the time, AND no error should be more than 10%."

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but is unable to accept or reject it at this time. We will consider this target in the context of future data releases and will report back to the committee.

25. RECOMMENDATION: The NAC recommends that the Census Bureau determine an accuracy target for differential privacy and children of color (Black, Hispanic, Asian, American Indian Alaska Native, and Native Hawaiian and Other Pacific Islander).

Census Bureau Response: The Census Bureau is evaluating this recommendation and asks the committee to provide additional feedback on what it thinks the accuracy targets for future data products should be.

Dr. William O'Hare's research on the Census Bureau's April 28th, 2021 demonstration product shows that DP creates over 91,000 Census blocks with children under age 18 but no adults. These implausible results are likely due to children being separated from their parents in differential privacy processing. Breaking the link produced absurd results that can impact the level of resources small communities get to serve children who live there, will lead to greater community distrust of the Census Bureau's data, and will likely affect subsequent Census products.

26. RECOMMENDATION: The NAC recommends that the Census Bureau implement differential privacy in a way that does not break the link between children and adults in a household.

Census Bureau Response: The Census Bureau accepts this recommendation. The Census Bureau recognizes the need for accurate data on children and on family composition for a wide range of data uses, which is why we designed the differentially private algorithms for the Detailed Demographic and Housing Characteristics (Detailed DHC) File to ensure that person-household joins will be preserved. We also designed a system for publication of the Demographic and Housing Characteristics (DHC) file to ensure that many tables that would have required a person-household join in the prior reporting system, such as proposed tables P17 "HOUSEHOLDS BY TYPE AND PRESENCE OF OWN CHILDREN UNDER 18" and P18 "HOUSEHOLDS BY PRESENCE OF PEOPLE 65 YEARS AND OVER, HOUSEHOLD SIZE, AND HOUSEHOLD TYPE," can be published alongside other DHC tables. For more information about the tables proposed for inclusion in the DHC and Detailed DHC, please visit our 2020 Census Data Products: Disclosure Avoidance Modernization page.

Impact on Tribal Nations

In its letter to the Interim Director dated May 21, 2021, the National Congress of American Indians (NCAI) identified the continued concerns of Tribal Nations regarding the impact of the application of the DAS to tribal geography and the 94-171 redistricting data set. The NAC joins NCAI in its concerns and the recommendations that NCAI made in its letter.

27. RECOMMENDATION: The NAC joins NCAI in recommending that the accuracy and usability of American Indian Alaska Native (AIAN) tribal geography data improve significantly in the final 2020 Census data, with any reductions in accuracy or worsening of error metrics being unacceptable.

Census Bureau Response: The Census Bureau accepts this recommendation. The 2020 Disclosure Avoidance System (DAS) production settings approved by the Data Stewardship Executive Policy (DSEP) Committee for the production run of the P.L. 94-171 Redistricting Data Summary File improved accuracy for AIAN tribal areas compared to the demonstration data.

28. RECOMMENDATION: The NAC joins NCAI in recommending that the Census Bureau ensure that the total population count in each AIAN tribal geography be held invariant (equal to the actual count), and that errors in other decennial demographic and housing characteristics must have no more than five percent error rates regardless of the size of the geography.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but must reject it. The list of invariants for the production run of the 2020 Census P.L. 94-171 Redistricting Data Summary File was finalized in November 2020. The Data Stewardship Executive Policy (DSEP) Committee did, however, make other changes to the

geographic spine and privacy-loss budget allocations to minimize the likelihood of distortions in the published data.

29. RECOMMENDATION: The NAC joins NCAI in recommending that the Census Bureau follow other standard practices of suppressing counts under 10 if there is a risk to privacy, but this should be done in consultation with the affected AIAN Tribal Nation(s), if applicable, and other affected groups.

Census Bureau Response: The Census Bureau thanks the NAC for this recommendation but must reject it. The Census Bureau's Data Stewardship Executive Policy (DSEP) Committee has already decided that all 2020 Census data products will be protected using the 2020 Census Disclosure Avoidance System. The 2020 Census Disclosure Avoidance System ensures that respondent confidentiality will be protected through a carefully calibrated noise infusion mechanism. As such, suppression of data is unnecessary to protect respondent confidentiality.